FACTORIZING PREDICTABLE STRUCTURE AND CONDITIONAL VARIATION FOR MULTIMODAL CARDIOVASCULAR SIGNAL GENERATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Multimodal cardiovascular signals, including electrocardiography (ECG), photoplethysmography (PPG), and arterial blood pressure (ABP), provide complementary observations of coupled cardiovascular dynamics. Generating a missing modality from partial observations is nevertheless underdetermined: deterministic predictors collapse unresolved variation into a single estimate, whereas direct conditional generators must model both observation-constrained structure and remaining variation with the same stochastic mechanism. We propose **CardioPRF**, a predictable–residual flow framework that factorizes these roles in latent space. A target-specific deterministic anchor first predicts target structure supported by the observed modalities, while global and high-resolution local conditions summarize cross-modal context and fine temporal evidence. Conditional flow matching then models only the residual around the anchor, and the target is decoded from their sum. A single model supports nine directed generation tasks across ECG, PPG, and ABP. Experiments show that deterministic prediction favors paired-target accuracy but has limited distributional coverage, whereas residual flow substantially improves distributional fidelity while retaining competitive reconstruction. Cross-dataset evaluations further indicate that generated signals preserve useful physiological information beyond the training dataset.

1 INTRODUCTION

Wearable and ambulatory sensing increasingly enables continuous cardiovascular monitoring outside episodic clinical measurements (Bayoumy et al., 2021). ECG measures cardiac electrical activity, PPG captures pulsatile peripheral blood-volume changes, and ABP reflects the resulting hemodynamic response. These signals are physiologically coupled but differ in acquisition burden and availability; PPG, in particular, is widely accessible in wearable devices (Charlton et al., 2022). This motivates cross-modal generation: recovering an unavailable cardiovascular waveform from modalities that are easier to observe.

Prior work has progressed from deterministic PPG-to-ABP reconstruction (Cheng et al., 2021; Ibtehaz et al., 2022), to adversarial PPG-to-ECG generation (Sarkar & Etemad, 2021), diffusion-based translation (Shome et al., 2024), latent rectified flow (Fang et al., 2025), and unified multimodal diffusion models (Chen et al., 2026). However, the conditional problem is generally underdetermined: an observed waveform constrains a missing modality without uniquely specifying its realization. Thus $p(x_t \mid X_{\mathcal{O}})$ contains both structure predictable from the observed subset $X_{\mathcal{O}}$ and variation that remains unresolved.

This distinction creates complementary limitations. Pointwise deterministic prediction tends toward a single conditional estimate and can suppress plausible variation. Direct stochastic generation can represent a distribution, but must regenerate both predictable structure and unresolved variation from the same stochastic process. We instead factorize the target latent as

$$z_t = z_{\mathrm{det}} + r, \tag{1}$$

where z_{det} is a deterministic anchor predicted from the observations and r is the residual modeled by conditional flow matching (Lipman et al., 2023). CardioPRF further uses a compact global condition for cross-modal context and a high-resolution local condition for waveform morphology.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

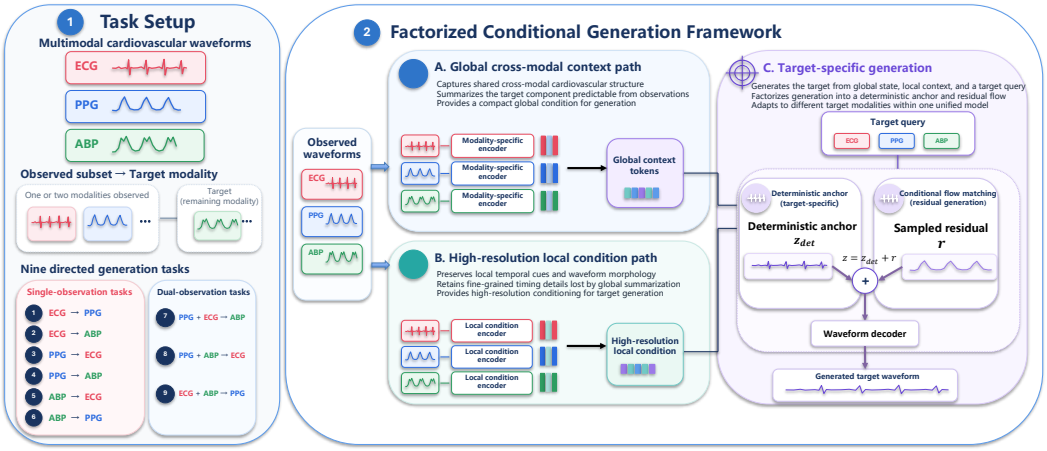

Figure 1: Overview of CardioPRF. Observed waveforms form complementary global and local conditions. A target-specific deterministic anchor predicts observation-constrained target structure, while conditional flow models only the remaining residual. The decoded anchor–residual sum forms the target waveform.

A single CardioPRF model supports all six single-observation and three dual-observation translations among ECG, PPG, and ABP. We evaluate paired-target accuracy, distributional fidelity, component ablations, and cross-dataset downstream utility. Our contributions are: (1) a predictable–residual formulation of underdetermined cardiovascular generation; (2) a target-specific deterministic anchor with conditional residual flow and complementary global/local conditioning; and (3) a unified nine-task evaluation showing that the factorization improves distributional fidelity while preserving competitive conditional and downstream performance.

2 Related Work

Cardiovascular cross-modal generation. ABP-Net (Cheng et al., 2021) and PPG2ABP (Ibtehaz et al., 2022) treat PPG-to-ABP reconstruction primarily as deterministic waveform prediction. For PPG-to-ECG, CardioGAN (Sarkar & Etemad, 2021) uses adversarial learning, RDDM (Shome et al., 2024) introduces region-disentangled diffusion, and PPGFlowECG (Fang et al., 2025) performs rectified-flow transport in an aligned latent space. These methods increase generative expressiveness, but the primary object remains the complete target waveform or latent. CardioPRF instead separates the observation- constrained component from the residual that requires stochastic modeling.

Unified and flow-based generation. UniCardio (Chen et al., 2026) demonstrates that heterogeneous cardiovascular denoising, imputation, and translation tasks can share a common diffusion framework. Our focus is complementary: we ask how the conditional target should be represented when only part of it is predictable from the available observations. More broadly, diffusion models learn iterative reverse dynamics (Ho et al., 2020), while Flow Matching trains continuous transport by regressing vector fields along conditional probability paths (Lipman et al., 2023); Rectified Flow encourages nearly straight transport trajectories (Liu et al., 2023). CardioPRF applies this transport only to the residual around an observation-conditioned anchor.

3 METHODOLOGY

3.1 PROBLEM FORMULATION AND CONDITIONING

Let $\mathcal{M} = \{\mathrm{ECG}, \mathrm{PPG}, \mathrm{ABP}\}$. For an observed subset $\mathcal{O} \subset \mathcal{M}$ of one or two modalities and target $t \in \mathcal{M} \setminus \mathcal{O}$, we model $p(x_t \mid X_{\mathcal{O}}, t)$. This yields six single-observation and three dual-observation directed tasks, all handled by one model.

Each observed modality is encoded in parallel. A modality-specific tokenizer produces global tokens, which are aggregated by learnable context queries using cross-attention and self-attention to form a compact global condition C_g. A separate multi-scale local encoder preserves high-resolution temporal evidence, yielding C_l. Local representations from different modalities are kept independent before concatenation, avoiding an assumption of pointwise cross-modal equivalence.

3.2 PREDICTABLE ANCHOR AND RESIDUAL FLOW

Each target modality uses an autoencoder $(\mathcal{E}_t, \mathcal{D}_t)$ with standardized latent z_t. A target-specific anchor predictor attends to both conditions and estimates

$$z_{\mathrm{det}} = A_t(C_g, C_l). \tag{2}$$

It is optimized by waveform reconstruction and latent regression,

$$\mathcal{L}_{\mathrm{det}} = \mathcal{L}_{\mathrm{wave}}(\mathcal{D}_t(z_{\mathrm{det}}), x_t) + \lambda_z \|z_{\mathrm{det}} - z_t\|_2^2. \tag{3}$$

After establishing the deterministic predictor, we define

$$r = z_t - \mathrm{sg}(z_{\mathrm{det}}), \tag{4}$$

where sg is stop-gradient. Conditional flow matching models $p(r \mid z_{\mathrm{det}}, C_g, C_l, q_t)$ with target query q_t. For $\epsilon \sim \mathcal{N}(0, I)$ and $\tau \sim \mathcal{U}(0, 1)$,

$$r_\tau = (1-\tau)\epsilon + \tau r, \qquad \mathcal{L}_{\mathrm{FM}} = \mathbb{E}\left[\|v_\theta(r_\tau, \tau \mid z_{\mathrm{det}}, C_g, C_l, q_t) - (r - \epsilon)\|_2^2\right]. \tag{5}$$

At inference, the learned ODE transports Gaussian noise to a residual $\hat{r}$, followed by $\hat{x}_t = \mathcal{D}_t(z_{\mathrm{det}} + \hat{r})$. Thus stochastic capacity is allocated to the unresolved component rather than to reconstructing the whole target latent.

3.3 STAGE-WISE OPTIMIZATION

We first pretrain modality-specific autoencoders, then train the deterministic anchor jointly over all nine tasks. The autoencoders and anchor are frozen while residual flow is learned from Eq. 4; a final low-rate joint refinement adapts the conditioning pathways, anchor, and flow while keeping the autoencoders fixed. This ordering gives the decomposition an explicit optimization semantics instead of learning an unconstrained $z_t = z_{\mathrm{det}} + r$ split from scratch.

4 EXPERIMENTS

4.1 EXPERIMENTAL SETUP

Datasets and preprocessing. We use MIMIC-BP (Sanches et al., 2024) for CardioPRF training, validation, main comparison, and ablations, while WESAD (Schmidt et al., 2018) and Cuffless BP (Kachuee et al., 2015) are reserved for external downstream evaluation. Table 1 gives the processed statistics. MIMIC-BP waveforms are resampled from 125 to 128 Hz. ECG and PPG are band-pass filtered at 0.5–40 and 0.5–8 Hz, respectively. Quality control is applied at both 30-s context and 4-s window levels; retained non-overlapping 4-s windows form the model inputs. ECG and PPG are normalized within their 30-s context, while model-input ABP is standardized using statistics computed only from clean training windows. External datasets are not used to train CardioPRF or select checkpoints.

Table 1: Processed datasets and experimental roles.

Dataset	Split	Subjects	Segments	Usage
MIMIC-BP	Train	971	128,258	Train
	Val.	176	23,288	Selection
	Test	204	27,019	Main/ablation
WESAD	Full	15	1,355	External
Cuffless BP	Full	–	336,382	External

Table 2: Main comparison. Lower is better. **Bold**/underline denote best/second-best within the same observation–target setting; rankings use unrounded values.

Target	Obs.	Method	MAE	Min-MAE	KS	FD
ECG	PPG	RDDM	0.38	0.34	0.70	31.66
		CardioGAN	0.34	0.26	**0.54**	4.80
		PPGFlowECG	**0.27**	**0.23**	0.59	6.16
		UniCardio	0.32	0.29	0.68	11.64
		CardioPRF	0.31	0.26	0.58	**1.32**
		CardioPRF-F	0.30	0.25	0.56	5.75
	PPG+ABP	CardioPRF-M	0.28	0.24	0.57	1.61
ABP	PPG	PPG2ABP	**0.47**	**0.42**	**0.33**	53.23
		UniCardio	0.92	0.85	0.51	78.19
		CardioPRF	0.60	0.54	0.40	**5.62**
		CardioPRF-F	0.59	0.53	0.40	6.90
	PPG+ECG	ABP-Net	0.63	0.59	**0.37**	297.09
		UniCardio	0.92	0.86	0.50	117.00
		CardioPRF-M	**0.54**	**0.49**	0.38	**8.18**

Implementation and baselines. Inputs are 4-s, 128-Hz windows (512 samples). CardioPRF uses hidden dimension 256, 32 global context tokens, five global attention blocks, eight heads, and dropout 0.1. The local encoder uses stride 4 and kernels $\{3, 7, 15, 31\}$; the anchor has three blocks. Autoencoders use eight latent channels with $16\times$ temporal downsampling. Residual flow uses dimension 256, five blocks, eight heads, and 64 local condition tokens per observed modality. Training uses 10/60/120/30 epochs for autoencoder, anchor, flow, and joint refinement stages, respectively; learning rate is 10^{-4} except 2×10^{-5} for refinement, batch size is 128, and weight decay is 10^{-4}. All nine tasks are sampled equally. Inference uses a 10-step Heun solver.

We retrain representative baselines on the same MIMIC-BP split and processed windows: CardioGAN (Sarkar & Etemad, 2021), RDDM (Shome et al., 2024), PPGFlowECG (Fang et al., 2025), UniCardio (Chen et al., 2026), PPG2ABP (Ibtehaz et al., 2022), and ABP-Net (Cheng et al., 2021). Architecture-specific objectives follow the corresponding methods, with model selection performed on validation data.

Metrics. We report MAE for paired-target error, Min-MAE as the best paired error among generated samples, Kolmogorov–Smirnov distance (KS), and raw Fréchet distance (FD) for distributional discrepancy. Lower is better for all four metrics. Downstream evaluations additionally report Macro-F1, HR MAE, and SBP/DBP MAE where applicable.

4.2 Main Comparison

Table 2 shows a consistent separation between paired-target and distributional objectives. For PPG→ECG, PPGFlowECG gives the lowest MAE/Min-MAE, whereas CardioPRF obtains the lowest FD (1.32). For PPG→ABP, PPG2ABP gives the lowest paired errors, while CardioPRF again substantially reduces FD (5.62 versus 53.23 and 78.19 for PPG2ABP and UniCardio). Adding a complementary observation improves paired-target prediction: CardioPRF-M reduces ECG MAE from 0.31 to 0.28 when ABP is added to PPG, and ABP MAE from 0.60 to 0.54 when ECG is

Table 3: Task-level generation-formulation ablation. Results are not averaged. Lower is better; **bold**/underline denote best/second-best.

Target	Obs.	Variant	MAE	Min-MAE	KS	FD
ABP	PPG	Anchor only	**0.49**	**0.44**	**0.38**	41.73
		Direct flow	0.59	0.53	0.40	12.94
		Anchor-residual	0.60	0.54	0.40	**5.62**
	PPG+ECG	Anchor only	**0.48**	**0.45**	**0.37**	24.39
		Direct flow	0.58	0.53	0.39	8.74
		Anchor-residual	0.58	0.53	0.40	**7.68**
ECG	PPG	Anchor only	**0.25**	**0.23**	0.62	26.63
		Direct flow	0.33	0.29	**0.58**	5.50
		Anchor-residual	0.31	0.26	0.58	**1.32**
	PPG+ABP	Anchor only	**0.24**	**0.22**	0.61	19.67
		Direct flow	0.30	0.27	**0.56**	2.98
		Anchor-residual	0.29	0.25	0.58	**1.36**

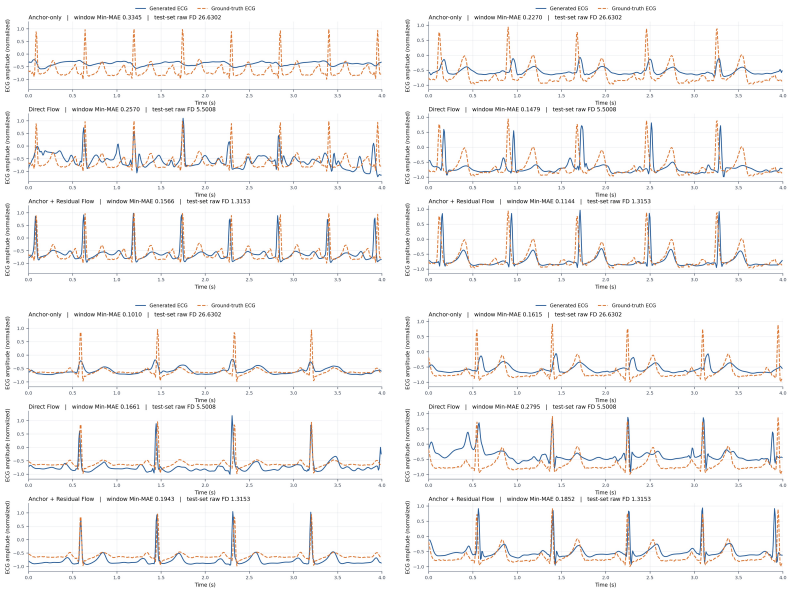

Figure 2: Representative PPG→ECG generations. Anchor-only outputs capture coarse rhythm but are often over-smoothed; direct flow restores amplitude yet can distort morphology or beat alignment; anchor-residual flow better preserves the rhythm scaffold while refining sharp morphology.

added. These results support evaluating reconstruction and distributional fidelity jointly rather than treating either as a complete measure of conditional generation quality.

4.3 Ablation Studies

Anchor-only prediction gives the lowest paired errors in all four tasks but much larger FD, consistent with a smooth conditional estimate that does not represent the full target distribution. Direct flow improves FD but must recover the entire target latent. Anchor-residual flow achieves the lowest FD in all four settings and improves paired errors over direct flow for both ECG tasks, supporting the proposed division of labor between deterministic prediction and stochastic residual generation.

Figure 2 illustrates the same trade-off. Over-smoothed anchor predictions can achieve favorable pointwise error while missing sharp ECG morphology. Direct flow produces sharper outputs but less stable local structure. Residual flow refines the deterministic scaffold and produces more realistic morphology, consistent with its substantially lower FD.

Table 4: Conditioning-pathway ablation. Values are unweighted macro-averages over PPG→ECG, PPG+ABP→ECG, PPG→ABP, and PPG+ECG→ABP.

Variant	MAE	Min-MAE	KS	FD
Full	<u>0.46</u>	<u>0.40</u>	<u>0.49</u>	**1.74**
w/o Global	0.47	0.42	0.49	44.60
w/o Local	**0.44**	**0.39**	**0.46**	<u>11.14</u>

Table 5: Cross-dataset downstream utility. WESAD: higher Macro-F1 and lower HR MAE are better. Cuffless BP: lower SBP/DBP MAE is better.

Dataset	Method	Metric 1	Metric 2
WESAD	RDDM	0.30	**11.55**
	PPGFlowECG	<u>0.30</u>	13.70
	UniCardio	0.28	37.90
	CardioPRF	**0.31**	<u>12.21</u>
Cuffless BP	UniCardio	40.42	26.04
	ABP-Net	31.97	10.04
	PPG2ABP	**19.15**	**8.02**
	CardioPRF	<u>19.18</u>	<u>9.32</u>

WESAD Metric 1/2: Macro-F1↑/HR MAE↓. Cuffless BP Metric 1/2: SBP MAE↓/DBP MAE↓. Rankings use unrounded values.

Table 4 shows complementary pathway roles. Removing the global condition raises FD from 1.74 to 44.60, indicating that compact cross-modal context strongly constrains the generated distribution. Removing the local condition slightly improves pointwise metrics but increases FD to 11.14, consistent with local temporal evidence being more important for realistic waveform variation than for minimizing a single paired-reference error. The full model therefore prioritizes distributional fidelity while retaining competitive pointwise accuracy.

4.4 Cross-Dataset Downstream Evaluation

WESAD and Cuffless BP are used only after MIMIC-BP training. We assess whether generated ECG retains information for the WESAD downstream classification protocol and heart-rate estimation, and whether generated ABP preserves SBP and DBP information on Cuffless BP.

On WESAD, CardioPRF gives the highest Macro-F1 (0.31) and second-lowest HR MAE (12.21), indicating that generated ECG retains useful information under dataset shift. On Cuffless BP, the task-specific PPG2ABP model is best, while CardioPRF is second for both SBP (19.18 versus 19.15) and DBP (9.32 versus 8.02), and improves substantially over UniCardio and ABP-Net. Because neither external dataset is used for CardioPRF training or checkpoint selection, these results provide evidence that the generated waveforms preserve transferable physiological information.

5 Conclusion

We presented CardioPRF, a unified framework for cardiovascular signal generation that separates observation-predictable target structure from unresolved conditional variation. A deterministic anchor provides a target-specific reference, while conditional flow models only the residual; global and local conditions supply complementary cross-modal and high-resolution information. Across ECG and ABP generation, this factorization consistently improves distributional fidelity relative to deterministic and direct-flow alternatives while retaining competitive paired-target accuracy. External WESAD and Cuffless BP results further support the physiological utility of the generated signals. A limitation is that the present evaluation focuses on ECG, PPG, and ABP and on the studied data distributions; broader clinical validation remains future work.

AI USE STATEMENT

Generative AI tools were used to assist with manuscript language editing and organization, literature-search support, and code debugging during model development. All AI-assisted text, code suggestions, citations, and analyses were reviewed and verified by the authors. Generative AI was not used to create experimental measurements or to alter reported results. The authors take full responsibility for the final manuscript, implementation, and reported findings.

REFERENCES

Karim Bayoumy, Mohammed Gaber, Abdallah Elshafeey, Omar Mhaimeed, Elizabeth H. Dineen, Francoise A. Marvel, Seth S. Martin, Evan D. Muse, Mintu P. Turakhia, Khaldoun G. Tarakji, and Mohamed B. Elshazly. Smart wearable devices in cardiovascular care: Where we are and how to move forward. *Nature Reviews Cardiology*, 18(8):581–599, 2021. doi: 10.1038/s41569-021-00522-7.

Peter H. Charlton, Panicos A. Kyriacou, Jonathan Mant, Vaidotas Marozas, Phil Chowienczyk, and Jordi Alastruey. Wearable photoplethysmography for cardiovascular monitoring. *Proceedings of the IEEE*, 110(3):355–381, 2022. doi: 10.1109/JPROC.2022.3149785.

Zehua Chen, Yuyang Miao, Liyuan Wang, Luyun Fan, Danilo P. Mandic, and Jun Zhu. Versatile cardiovascular signal generation with a unified diffusion transformer. *Nature Machine Intelligence*, 8:6–19, 2026. doi: 10.1038/s42256-025-01147-y.

Juan Cheng, Yufei Xu, Rencheng Song, Yu Liu, Chang Li, and Xun Chen. Prediction of arterial blood pressure waveforms from photoplethysmogram signals via fully convolutional neural networks. *Computers in Biology and Medicine*, 138:104877, 2021. doi: 10.1016/j.compbiomed.2021.104877.

Xiaocheng Fang, Jiarui Jin, Haoyu Wang, Che Liu, Jieyi Cai, Guangkun Nie, Jun Li, Hongyan Li, and Shenda Hong. PPGFlowECG: Latent rectified flow with cross-modal encoding for PPG-guided ECG generation and cardiovascular disease detection. *arXiv preprint arXiv:2509.19774*, 2025.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *Advances in Neural Information Processing Systems*, volume 33, pp. 6840–6851, 2020.

Nabil Ibtehaz, Sakib Mahmud, Muhammad E. H. Chowdhury, Amith Khandakar, Muhammad Salman Khan, Mohamed Arselene Ayari, Anas M. Tahir, and M. Sohel Rahman. PPG2ABP: Translating photoplethysmogram (PPG) signals to arterial blood pressure (ABP) waveforms. *Bioengineering*, 9(11):692, 2022. doi: 10.3390/bioengineering9110692.

Mohamad Kachuee, Mohammad Mahdi Kiani, Hoda Mohammadzade, and Mahdi Shabany. Cuffless high-accuracy calibration-free blood pressure estimation using pulse transit time. In *2015 IEEE International Symposium on Circuits and Systems (ISCAS)*, pp. 1006–1009, 2015. doi: 10.1109/ISCAS.2015.7168806.

Yaron Lipman, Ricky T. Q. Chen, Heli Ben-Hamu, Maximilian Nickel, and Matthew Le. Flow matching for generative modeling. In *International Conference on Learning Representations*, 2023.

Xingchao Liu, Chengyue Gong, and Qiang Liu. Flow straight and fast: Learning to generate and transfer data with rectified flow. In *International Conference on Learning Representations*, 2023.

Ivandro Sanches, Victor V. Gomes, Carlos Caetano, Lizeth S. B. Cabrera, Vinicius H. Cene, Thomas Beltrame, Wonkyu Lee, Sanghyun Baek, and Otávio A. B. Penatti. MIMIC-BP: A curated dataset for blood pressure estimation. *Scientific Data*, 11:1233, 2024. doi: 10.1038/s41597-024-04041-1.

Pritam Sarkar and Ali Etemad. CardioGAN: Attentive generative adversarial network with dual discriminators for synthesis of ECG from PPG. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pp. 488–496, 2021. doi: 10.1609/aaai.v35i1.16126.

Philip Schmidt, Attila Reiss, Robert Duerichen, Claus Marberger, and Kristof Van Laerhoven. Introducing WESAD, a multimodal dataset for wearable stress and affect detection. In *Proceedings of the 20th ACM International Conference on Multimodal Interaction*, pp. 400–408, 2018. doi: 10.1145/3242969.3242985.

Debaditya Shome, Pritam Sarkar, and Ali Etemad. Region-disentangled diffusion model for high-fidelity PPG-to-ECG translation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pp. 15009–15019, 2024. doi: 10.1609/aaai.v38i13.29422.